Refinement of the Solution Structure of the DNA Hexamer 5′d(GCATGC)$_2$: Combined Use of Nuclear Magnetic Resonance and Restrained Molecular Dynamics[†]

Michael Nilges, G. Marius Clore,* and Angela M. Gronenborn
Max-Planck-Institut für Biochemie, D-8033 Martinsried bei München, FRG

Axel T. Brunger and Martin Karplus
Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

Lennart Nilsson
Department of Medical Biophysics, Karolinska Institute, S-10401 Stockholm, Sweden

Received January 26, 1987; Revised Manuscript Received March 24, 1987

ABSTRACT: The solution structure of the self-complementary DNA hexamer 5′d(GCATGC)$_2$ comprising the specific target site for the restriction endonuclease *Sph*I is investigated by using nuclear magnetic resonance spectroscopy and restrained molecular dynamics. All the nonexchangeable proton resonances are assigned sequentially, and from time-dependent nuclear Overhauser enhancement measurements a set of 158 approximate interproton distances are determined. These distances are used as the basis of a structure refinement using restrained molecular dynamics in which the interproton distances are incorporated into the total energy function of the system in the form of an effective potential term. Two restrained molecular dynamics simulations are carried out, starting from classical B- and A-DNA [atomic root mean square (rms) difference 3.3 Å]. In both cases convergence is achieved to essentially identical structures satisfying the experimental restraints and having a root mean square difference of only 0.3 Å between them, which is within the rms fluctuations of the atoms about their average positions. These results suggest that the restrained molecular dynamics structures represent reasonable approximations of the solution structure. The converged structures are of the B type and exhibit clear sequence-dependent variations of helical parameters, some of which follow Calladine's rules and can be attributed to the relief of interstrand purine–purine clash at adjacent base pairs. In addition, the converged restrained dynamics structures appear bent with a radius of curvature of approximately 20 Å. This bending appears to be due almost entirely to the large positive base roll angles, particularly at the Pyr-Pur steps. Further, the global and local helix axes are not coincident, and the global helix axis represents a superhelical axis which the bent DNA, when extended into an "infinite" helix by repeated translation and rotation, wraps around.

X-ray crystallographic studies have shown that the structures of DNA oligonucleotides in crystals are not regular helices but exhibit local structural variations that appear to be sequence dependent (Dickerson & Drew, 1981; Dickerson et al., 1983; Wang et al., 1983; Shakked et al., 1983; McCall et al., 1985). It has been suggested that these structural features may play an important role in both specific and nonspecific DNA–protein interactions (Lomonossoff et al., 1981; Rhodes, 1982; Drew & Travers, 1984, 1985; Frederick et al., 1984). A knowledge of the structural details of oligonucleotides in solution is therefore desirable. The method of choice is nuclear magnetic resonance (NMR)[1] spectroscopy and, in particular, nuclear Overhauser enhancement (NOE) measurements to demonstrate the proximity of protons in space and to determine their separation (Noggle & Schirmer, 1971). These interproton distances can then be used as the basis of a structure determination or refinement. The method we have chosen is restrained molecular dynamics in which the interproton distances are incorporated into the total energy function of the system in the form of effective potentials and energetic considerations are taken into account during the entire course of the structure refinement or determination (Kaptein et al., 1985; Clore et al., 1985). This particular method has been shown to have a large radius of convergence when applied to both proteins (Brunger et al., 1986; Clore et al., 1986a) and oligonucleotides (Nilsson et al., 1986). Alternative approaches that also have large radii of convergence include metric matrix distance geometry calculations (Crippen & Havel, 1978; Havel & Wüthrich, 1985) and restrained least-squares refinement in torsion angle space using variable target functions (Braun & Go, 1985). To date, the former has been applied to both proteins (Williamson et al., 1985) and oligonucleotides (Hare & Reid, 1986), while the latter has only been applied to proteins (Kline et al., 1986).

In this paper we focus our attention on the solution structure of the alternating Pur-Pyr DNA oligonucleotide 5′d-(GCATGC)$_2$ comprising the specific target site for the restriction endonuclease *Sph*I. First, all the nonexchangeable proton resonances are assigned in a sequential manner. From time-dependent NOE measurements a set of approximate interproton distances is then derived which is used as the basis of a structure refinement by restrained molecular dynamics.

[†] This work was supported by the Max-Planck Gesellschaft and Grant Cl 86/1-1 of the Deutsche Forschungsgemeinschaft (G.M.C. and A.M.G.). M.N. thanks the Max-Planck Gesellschaft for a Max-Planck predoctoral fellowship.

[1] Abbreviations: NMR, nuclear magnetic resonance spectroscopy; NOE, nuclear Overhauser enhancement; NOESY, two-dimensional NOE spectroscopy; HOHAHA, two-dimensional homonuclear Hartmann–Hahn spectroscopy; rms, root mean square; RD, restrained dynamics; FD, free dynamics; EDTA, ethylenediaminetetraacetic acid; 2D, two dimensional.

0006-2960/87/0426-3718$01.50/0 © 1987 American Chemical Society

As in our earlier study (Nilsson et al., 1986), convergence is achieved by starting from two quite different initial structures, namely, classical A- and B-DNA. The atomic rms difference between the average restrained dynamics structures is only 0.3 Å, compared to a value of 3.2 Å between the two initial structures, and is less than the magnitude of the rms atomic fluctuations of either refined structure. The converged structures are of the B type and exhibit sequence-dependent variations of helical parameters, some of which are in accordance with Calladine's predictions (Calladine, 1982; Dickerson, 1983; Calladine & Drew, 1984). In addition, the converged structures possess an unusual structural feature: namely, they appear to be bent. The bending can be attributed almost entirely to the large positive base roll angles, particularly at the Pyr-Pur steps. Further, the global and local helix axes are not coincident, and the global helix axis is seen to represent a superhelical axis which the bent DNA wraps around.

Materials and Methods

The DNA hexamer 5′d(GCATCG)$_2$ was prepared from suitably protected nucleosides according to the phosphite triester method (Matteuci & Caruthers, 1981) principally as described by Seliger et al. (1982) and purified by reverse-phase high-pressure liquid chromatography using a Walters μBondapak C_{18} column. After desalting and extensive lyophilization, the hexamer (final concentration 4 mM) was taken up in 99.96% D_2O containing 500 mM KCl, 50 mM potassium phosphate, pH* 6.8 (meter reading uncorrected for the isotope effect on the glass electrode), and 0.02 mM EDTA.

The temperature used for all NMR experiments was 20 °C. Under these conditions of ionic strength and temperature, the hexamer was entirely double stranded as judged from thermal denaturation studies (unpublished data). It was also B type as judged from its circular dichroism spectrum (unpublished data).

All NMR spectra were recorded on a Bruker AM500 spectrometer equipped with an ASPECT 3000 computer and digital phase shifters. Quadrature detection was used with the carrier placed at the position of the residual HOD resonance, approximately in the middle of the nucleic acid spectrum. Chemical shifts are expressed relative to sodium 4,4-dimethyl-4-silapentane-1-sulfonate.

Through-bond (direct and relayed) and through-space connectivities were determined by pure-phase absorption two-dimensional MLEV17 Hartmann-Hahn (Davis & Bax, 1985; Bax & Davis, 1985) and NOESY (Jeener et al., 1979) spectroscopy, respectively. The pure-phase absorption spectra were recorded by using the time-proportional phase incrementation method (Redfield & Kuntz, 1975; Bodenhausen et al., 1980) as described by Marion and Wüthrich (1983). Appropriate phase cycling was used to eliminate axial peaks and, in the case of the NOESY spectra, peaks due to multiple quantum coherence transfer. For both types of 2D spectra 128 transients were collected for each of 512 increments with a relaxation delay of 1 s between successive transients. The sweep width employed was 6042 Hz. A square 1K × 1K frequency domain matrix was obtained by zero filling in the t_1 dimension to give a digital resolution of 5.88 Hz per point in both dimensions. An initial phase correction was carried out during transformation with a final adjustment after completion of the two-dimensional transform. These manipulations were followed by symmetrization (Baumann et al., 1981).

One-dimensional NOE spectra were recorded with a 90° observation pulse, an acquisition time of 0.5 s, and a relaxation delay of 2 s. The NOEs were observed by directly collecting the difference free induction decay by interleaving eight transients after saturation for a set time of a given resonance with eight transients of off-resonance irradiation (applied for the same length of time), negating the memory between eight transient cycles. The irradiation power used was sufficient to be in the high-power limit so that saturation was effectively instantaneous while selectivity was preserved so that only a single resonance at a time was saturated (Dobson et al., 1982). A total of 1600 transients were recorded for each difference spectrum, and prior to Fourier transformation the difference free induction decays were multiplied by an exponential equivalent to a line broadening of 2 Hz.

NOE intensities were determined by peak integration in the case of the one-dimensional measurements and by two-dimensional integration of cross-peaks in the case of the two-dimensional experiments.

All energy minimization and molecular dynamics calculations were carried out by using the program CHARMM (Brooks et al., 1983) optimized for the CRAY computer (Brunger, unpublished data). The empirical energy function used was that developed by Nilsson and Karplus (1985) for nucleic acids in which all hydrogen atoms are treated explicitly. The effective potential E_{NOE} representing the interproton distance restraints was added to the total energy function of the system in the form of a skewed biharmonic effective potential (Clore et al., 1985) given by

$$\begin{aligned} E_{\mathrm{NOE}}(r_{ij}) &= c_1(r_{ij} - r_{ij}^{\circ})^2 \quad \text{if } r_{ij} > r_{ij}^{\circ} \\ E_{\mathrm{NOE}}(r_{ij}) &= c_2(r_{ij} - r_{ij}^{\circ})^2 \quad \text{if } r_{ij} < r_{ij}^{\circ} \end{aligned} \tag{1}$$

where r_{ij}° and r_{ij} are the target and calculated interproton distances, respectively, and c_1 and c_2 are force constants given by

$$c_1 = \frac{k_B T S}{2(\Delta_{ij}^{+})^2} \qquad c_2 = \frac{k_B T S}{2(\Delta_{ij}^{-})^2} \tag{2}$$

where k_B is the Boltzmann constant, T is the absolute temperature, S is a scale factor, and Δ_{ij}^{+} and Δ_{ij}^{-} are the positive and negative error limits on the value of r_{ij}. Solvent molecules were not included explicitly in the calculations, but the effect of solvent was approximated by a $1/r$ screening function (Gelin & Karplus, 1977; Brooks et al., 1983) and by reducing the net charge on the phosphate group to $-0.32e$ (Tidor et al., 1982). The nonbonded interactions were switched off, by using a cubic switching function, between 9.5 and 10.5 Å, with pairs up to 11.5 Å included in the nonbonded list. Integration of the classical equations of motion was performed by use of a Verlet integration algorithm (Verlet, 1967) with initial velocities assigned from a Maxwellian distribution at 300 K. The time step of the integrator was 0.001 ps, and the nonbonded interaction lists were updated every 0.02 ps. Bond lengths involving hydrogen atoms were kept fixed with the SHAKE algorithm (Ryckaert et al., 1977).

Displaying of trajectories was carried out on an Evans & Sutherland PS330 color graphics system using a modified version of the function network of FRODO (Jones, 1978, 1982) interfaced with CHARMM. Analysis of helical parameters was carried out by using modified versions of the AHELIX (written by J. Rosenberg) and BROLL and CYLIN (written by R. E. Dickerson) programs adapted to deal with dynamics trajectories.

Results and Discussion

Assignment of Proton Resonances and Low-Resolution Structure. The proton resonances of the hexamer were as-

Table I: Proton Resonance Assignments of the Hexamer at 20 °C

	chemical shift (ppm)[a]									
residue	H8/H6	H5/CH_3	H2	H1′	H2′	H2″	H3′	H4′	H5′	H5″
G_1	7.98			5.97	2.62	2.77	4.84	4.25	3.72	3.72
C_2	7.49	5.45		5.68	2.16	2.49	4.90	4.21	4.13	4.13
A_3	8.37		7.72	6.30	2.74	2.96	5.06	4.44	4.10	4.19
T_4	7.13	1.49		5.76	1.96	2.35	4.87	4.17	4.28	4.15
G_5	7.86			5.92	2.60	2.69	4.98	4.37	4.12	4.09
C_6	7.43	5.37		6.18	2.19	2.19	4.49	4.06	4.25	4.23

[a] The chemical shifts are expressed relative to sodium 4,4-dimethyl-4-silapentane-1-sulfonate.

signed in a sequential manner (Reid et al., 1983; Scheek et al., 1983; Hare et al., 1983; Feigon et al., 1983; Clore & Gronenborn, 1983; Clore et al., 1984; Weiss et al., 1984). This involved (a) the use of Hartmann–Hahn spectroscopy to demonstrate direct and relayed through-bond connectivities along the H1′ ↔ H2′/H2″ ↔ H3′ ↔ H4′ ↔ H5′/H5″ pathway within each sugar unit and (b) the use of NOESY spectroscopy to demonstrate through-space (<5-Å) connectivities along the H1′(i − 1) ↔ H8/H6(i) ↔ H1′(i), H2′/H2″(i − 1) ↔ H8/H6(i) ↔ H2′/H2″(i) and H8/H6(i) ↔ H5/CH_3(i + 1) pathways. Examples of various regions of the Hartmann–Hahn and NOESY spectra are shown in Figure 1. The distinction between the H2′ and H2″ protons was easily made from the relative NOE cross-peak intensities of the intranucleotide H1′–H2′ and H1′–H2″ NOEs as the distance between the H1′ and H2″ protons is usually shorter and can never be longer than that between the H1′ and H2′ protons for all sugar pucker conformations. The H5′ and H5″ protons could also be distinguished as in right-handed DNA the H1′ proton of residue i is much closer to the H5′ proton than to the H5″ proton of residue i + 1. The complete list of assignments is given in Table I.

The low-resolution solution structure of an oligonucleotide is readily deduced from a qualitative assessment of the relative NOE cross-peak intensities [for a review see Clore and Gronenborn (1985a) and Gronenborn and Clore (1985)] as these are approximately proportional to $\langle r^{-6}\rangle$ at short mixing times. In the case of the hexamer, the pattern of NOE cross-peak intensities is indicative of a right-handed B-type structure. The handedness is ascertained from the observation of H8/H6(i)–H5/CH_3(i + 1) NOEs. The overall B-type structure is established from the observation of the following pattern of NOE intensities involving the H2′,H2″ and H8/H6 protons: $N_{H2'(i)-H8/H6(i)} \gg N_{H2''(i-1)-H8/H6(i)} > N_{H2'(i-1)-H8/H6(i)}$. Finally, the sugar pucker and glycosidic bond conformations lie in the O1′-endo to C2′-endo and anti ranges, respectively, on account of the $N_{H2'(i)-H8/H6(i)} \gg N_{H1'(i)-H8/H6(i)} > N_{H3'(i)-H8/H6(i)}$ pattern of intranucleotide sugar–base NOEs.

Interproton Distances. In order to determine interproton distances, cross-relaxation rates were measured from the time dependence of the NOEs as the initial buildup rate of the NOE is equal to the cross-relaxation rate (Wagner & Wüthrich, 1979; Dobson et al., 1982; Clore & Gronenborn, 1985b). For this purpose we principally used one-dimensional NOE measurements, irradiating every resonance in turn, as in this particular case we found their quantification to be more reliable than for the two-dimensional measurements (viz., relative errors of ≲±15% for the former compared to ±20–30% for the latter). Some examples of NOE time courses are shown in Figure 2. $(\langle r^{-6}\rangle)^{-1/6}$ mean interproton distances were then obtained from

$$(\langle r_{ij}^{-6}\rangle)^{-1/6} = (\sigma_{kl}/\sigma_{ij})^{1/6} r_{kl} \qquad (3)$$

where r_{ij} is the unknown distance, r_{kl} the appropriate fixed internal reference distance, and σ_{kl} the corresponding cross-relaxation rates derived from the initial slopes of the time-dependent NOEs (Wagner & Wüthrich, 1979; Dobson et al., 1982; Clore & Gronenborn, 1985b). We note, however, that the ratios of the NOESY cross-peak intensities from the 100 ms NOESY spectrum were essentially identical with the corresponding ratios of initial slopes derived from the one-dimensional measurements, within the errors expected for two-dimensional integration (±20–30%). In the case of oligonucleotides there are three suitable reference distances whose values are fixed by the geometry of the sugars and bases themselves: namely, $r_{H2'-H2''}$, $r_{C(H5)-C(H6)}$, and $(\langle r_{T(CH3)-T(H6)}\rangle^{-6})^{-1/6}$, which have values of 1.8, 2.5, and 2.7 Å, respectively. Equation 3 is only valid if the effective correlation times of the i–j and k–l interproton vectors are approximately the same. The effective correlation time τ_{eff} of the fixed-distance interproton vectors is simply obtained from (Solomon, 1955)

$$\sigma_{ij} = \frac{\hbar^2\gamma^4}{10 r_{ij}^6}\left(\frac{6\tau_{eff}}{1 + 4\omega^2\tau_{eff}^2} - \tau_{eff}\right) \qquad (4)$$

where $\hbar$ is Planck's constant divided by 2π, γ the gyromagnetic ratio of the proton, and ω the spectrometer frequency. In this respect, we could observe *no* significant residue to residue variation in the effective correlation times of the three fixed-distance interproton vectors. As observed previously (Gronenborn et al., 1984a; Clore & Gronenborn, 1984), however, the effective correlation times of the H2′–H2″ vectors (~1 ns) were significantly shorter than those of the C(H5)–C(H6) and T(CH_3)–T(H6)vectors (~2.5 ns). Similar conclusions have been arrived at by using ^{13}C and ^{31}P NMR (Bolton & James, 1979, 1980; Hogan & Jardetzky, 1979). Consequently, the appropriate choice of reference distance in the calculation of the unknown interproton distances has to be made. This has been discussed and verified in detail previously (Gronenborn et al., 1984a; Gronenborn & Clore, 1985) so only the conclusions will be given here: namely, all unknown distances involving sugar–sugar and sugar–base (with the exception of the H1′ sugar–base) vectors should be calculated by using the H2′–H2″ cross-relaxation rate and distance as a reference, while all those involving base–base and sugar H1′–base vectors should be calculated by using the C(H5)–C(H6) or T(CH_3)–T(H6) cross-relaxation rates and distances as a reference. This choice is based on the reasoning that the contribution from internal motion to the effective correlation times of the first class of distances will mainly be dominated by motion within the sugar units whereas that of the second class will mainly be dominated by motion about the glycosidic bond. [Note that rotation of the methyl protons makes an insignificant contribution to the effective correlation time of the T(CH_3)–H vectors; see Keepers and James (1984), Clore and Gronenborn (1984), and Gronenborn and Clore (1985).] The validity of this reasoning is easily checked by calculating distances, which, although not fixed, have a very limited range (±0.2 Å) of values [e.g., the intrasugar H1′–H2″ and H2′–H3′

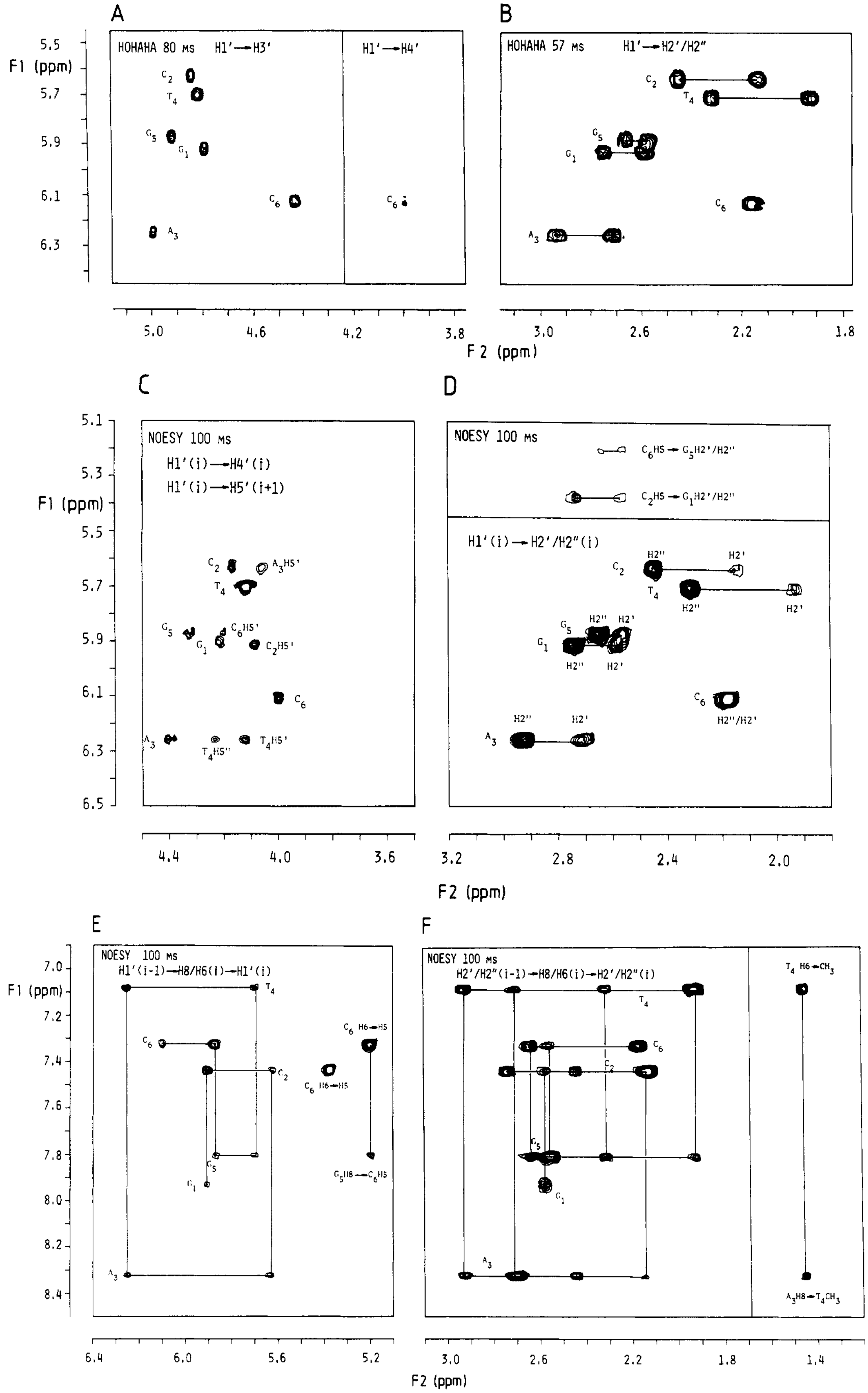


FIGURE 1: Pure-phase absorption NOESY and homonuclear Hartmann–Hahn (HOHAHA) spectra of 5'd(GCATCG)$_2$ (A) and (B) show Hartmann–Hahn spectra (mixing times 57 and 80 ms) of the H1'(*F*1 axis)–H3'/H4'(*F*2 axis) and H1'(*F*1 axis)–H2'/H2"(*F*2 axis) regions, respectively. The cross-peaks in (A) arise from direct through-bond connectivities whereas those in (B) arise from relayed through-bond connectivities. (C), (D), (E), and (F) show NOESY spectra (mixing time 100 ms) of the H1'(*F*1 axis)–H4'/H5"(*F*2 axis), H1'/H3'(*F*1 axis)–H2'/H2"(*F*2 axis), H8/H6(*F*1 axis)–H1'/H5(*F*2 axis), and H8/H6(*F*1 axis)–H2'/H2"/CH$_3$(*F*2 axis), respectively. The cross-peaks in (C)–(F) arise from through-space connectivities between protons separated by <5 Å.

Table II: Cross-Relaxation Rates for the Hexamer Determined from Time-Dependent NOE Measurements together with the $(\langle r^{-6} \rangle)^{-1/6}$ Mean Interproton Distances Calculated from Them[a]

Intranucleotide

	residue											
	G_1		C_2		A_3		T_4		G_5		C_6	
proton pair	σ_{ij} (s^{-1})	r_{ij} (Å)	σ_{ij} (s^{-1})	r_{ij} (Å)	σ_{ij} (s^{-1})	r_{ij} (Å)	σ_{ij} (s^{-1})	r_{ij} (Å)	σ_{ij} (s^{-1})	r_{ij} (Å)	σ_{ij} (s^{-1})	r_{ij} (Å)
sugar–sugar												
H1–H2′	0.21	2.5	0.17	2.6	0.24	2.4	0.26	2.4	0.19	2.5		
H1–H2″	0.38	2.3	0.45	2.2	0.46	2.2	0.50	2.1	0.32	2.3		
H1′–H4′	0.10	2.8	0.13	2.7	0.11	2.8	0.14	2.7	0.11	2.8	0.14	2.7
H2′–H2″	1.50	1.8	1.45	1.8	1.40	1.8	1.45	1.8				
H2′–H3′	0.36	2.3	0.47	2.2	0.39	2.2	0.24	2.4	0.37	2.3		
H2″–H3′	0.21	2.5	0.22	2.5	0.27	2.4	0.10	2.8	0.25	2.4		
H3′–H4′	0.21	2.5	0.13	2.7	0.24	2.4	0.10	2.8	0.19	2.5		
H3′–H5″					0.36	2.3						
sugar–base												
H1′–H8/H6	0.07	3.6	0.10	3.4	0.10	3.4	0.08	3.5	0.10	3.4	0.13	3.2
H2′–H8/H6	0.44	2.2	0.68	2.0	0.61	2.1	0.70	2.0	0.48	2.2	0.50	2.1
H3′–H8/H6	nd		lag	>3.5	lag	>3.5	lag	>3.5	lag	>3.5	nd	
base–base												
H6–H5			0.65	2.5							0.60	2.5
H6–CH_3							0.40	2.7				

Internucleotide

		base step									
proton of	proton of	G_1pC_2		C_2pA_3		A_3pT_4		T_4pG_5		G_5pC_6	
5′-residue	3′-residue	σ_{ij} (s^{-1})	r_{ij} (Å)	σ_{ij} (s^{-1})	r_{ij} (Å)	σ_{ij} (s^{-1})	r_{ij} (Å)	σ_{ij} (s^{-1})	r_{ij} (Å)	σ_{ij} (s^{-1})	r_{ij} (Å)
H1′	H8/H6	0.14	3.2	0.09	3.5	0.14	3.2	0.06	3.9	0.18	3.0
H2′	H8/H6	0.10	2.8	0.14	2.7	0.13	2.7	0.17	2.6	0.11	2.8
H2″	H8/H6	0.25	2.5	0.19	2.5	0.34	2.3	0.22	2.5	0.24	2.4
H8/H6	H8/H6	nd		0.04	4.0	0.06	3.7	0.10	3.4	0.06	3.7
H8/H6	H5/CH3	nd				0.29	2.8			0.08	3.5
H1′	H5/CH3	0.10	3.4			0.19	3.0			0.07	3.6
H2′	H5/CH3	0.10	2.8			0.17	2.6			0.09	2.9
H2″	H5/CH3	0.15	2.6			0.23	2.4			0.10	2.8
H2	H1′					0.06	3.8				

[a] When the interproton distances were calculated from the cross-relaxation rates by using eq 3, the H2′–H2″ cross-relaxation rate and distance were used for all sugar–sugar and sugar–base (with the exception of sugar H1′–base) distances, and the C(H5)–C(H6) cross-relaxation rate and distance was used for all base–base and H1′–sugar base distances (see text). The estimated errors in the distances are as follows: for distances calculated by using the H2′–H2″ vector as a reference they are −0.2/+0.3 Å for $r < 2.5$ Å and −0.3/+0.4 Å for $r \geq 2.5$ Å; and for distances calculated by using the C(H5)–C(H6) vector as a reference they are −0.2 Å/+0.3 Å for $r < 3$ Å and −0.3 Å/+0.4 Å for $r \geq 3$ Å. Symbol: nd, not detectable.

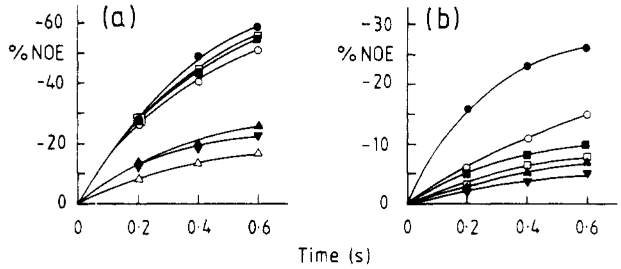


FIGURE 2: Time dependence of some NOEs measured from one-dimensional spectra. The time dependences of NOEs between the H2′ and H2″ protons, between the C(H5)–C(H6) protons, and between the $T(CH_3)$–T(H6) protons are shown in (a). These NOEs constitute internal references as they involve protons a fixed distance apart. The time dependences of NOEs observed on irradiation of the A_3(H8) resonance are shown in (B). The symbols are as follows. In (a): ●, G_1(H2′)–G_1(H2″); □, C_2(H2′)–C2(H2″); ○, A_3(H2′)–A_3(H2″); ■, T_4(H2′)–T_4(H2″); ▲, C_2(H5)–C_2(H6); ▼, C_6(H5)–C_6(H6); △, $T_4(CH_3)$–T_4(H6). In (b): ●, A_3(H8)–A_3(H2′); ○, A_3(H8)–$T_4(CH_3)$; ■, A_3(H8)–C_2(H2″); □, A_3(H8)–C_2(H2′); ▲, A_3(H8)–A_3(H1′); ▼, A_3(H8)–C_2(H1′).

distances and the intra base pair T(H3)–A(H2) distance].

In addition to the random experimental errors which are expected to be Gaussian in distribution, the use of eq 3 also introduces a small systematic error due to the presence of a small amount of unavoidable spin diffusion. This has been analyzed by Clore and Gronenborn (1985b), who showed that the unknown distance will become closer to that of the reference distances as the extent of spin diffusion increases. In the case of the present data, the unknown distances are all larger than the corresponding reference distances so that their values will tend to be systematically underestimated. The errors in the estimated values of the unknown distances are therefore skewed with the size of the systematic error, depending on the value of the unknown distance relative to that of the reference distance. The smaller the value of the unknown distance, the smaller will be the contribution of spin diffusion to the measured cross-relaxation rate and, consequently, the smaller the error [for a discussion of these effects see also Nilsson et al. (1986)]. On the basis of our previous calculations (Clore & Gronenborn, 1985b), we have estimated the errors as follows: for distances calculated by using the H2′–H2″ as a reference, the errors are −0.2/+0.3 Å and −0.3/+0.4 Å for $r_{ij} < 2.5$ Å and $r_{ij} \geq 2.5$ Å, respectively; for distances calculated by using the C(H5)–C(H6) or T-(CH_3)–T(H6) vector as a reference, they are −0.2/+0.3 Å and −0.3/+0.4 Å for $r_{ij} < 3.0$ Å and $r_{ij} \geq 3.0$ Å, respectively. In making these error estimates, we have been slightly more conservative than in our previous restrained dynamics study on the DNA oligonucleotide 5′d$(CGTACG)_2$ (Nilsson et al., 1986). Further, they are sufficiently generous to ensure that errors arising from variations in effective correlation times are minimal.

A summary of the measured cross-relaxation rates together with the distances calculated from them is given in Table II.

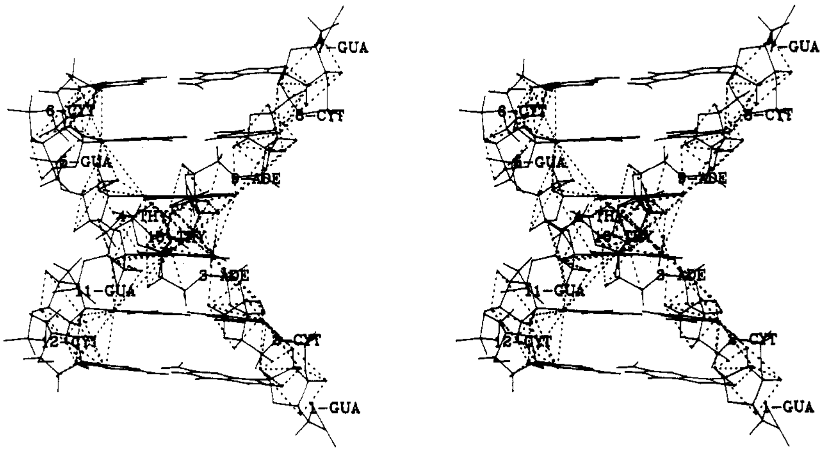


FIGURE 3: Stereoview of the interproton distance restraints as dashed lines on a classical B-DNA framework. Note that the values of the distance restraints in these figures are those found in classical B-DNA and *not* the experimental values.

A stereoview of this distance set, comprising 158 distances, superimposed on a classical B-DNA framework is shown in Figure 3.

Structure Refinement. In order to obtain an approximate picture of the structure of the hexamer in solution, we proceeded to carry out restrained molecular dynamics calculations, incorporating the experimental interproton distances into the total energy function of the system in the form of effective potentials (see eq 1). Two refinement calculations were carried out, starting from two quite different structures, namely, classical B- (IniI) and A- (IniII) DNA (see Figure 4). Each structure was then subjected to the following steps: (i) 500 cycles of restrained energy minimization with the restraints scale factor set to 0.25 (see eq 2); (ii) 1 ps of equilibration during which time the structure was heated up from 200 to 300 K in steps of 10 K every 0.1 ps and S was increased from 0.25 to 2.75 in steps of 0.25 every 0.1 ps; (iii) 15 ps of restrained dynamics (known as the first dynamics run) with the initial velocities assigned at 300 K and S set to 3; (iv) 5 ps of thermalization with S set to 3 during which time the velocities were reassigned every 0.2 ps at 300 K; and (v) 30 ps of restrained dynamics (known as the second dynamics run) with the initial velocities assigned at 300 K and S set to 3. The average restrained dynamics structures RDI and RDII were then obtained by averaging the coordinate trajectories over the entire 30 ps of the second dynamics run. Two other sets of calculations were carried out for comparison. In the first set, the initial structures were subjected to 2000 cycles of restrained energy minimization with S set to 3, and this set resulted in the structure known as RMI and RMII. In the second set, free dynamics simulations (i.e., with no interproton distance restraints and S set to 0) were carried out starting from the two initial structures. These involved the same steps as the restrained dynamics simulations except that the calculations ended after the first 15-ps dynamics run. The two average free dynamics structures, known as FDI and FDII, were obtained by averaging the coordinate trajectories over the last 10 ps of the dynamics run.

The atomic rms differences between the structures are given in Table III, the rms differences between the calculated and experimental interproton distances in Table IV, and the energies of the various structures in Table V. Stereoviews of the initial structures and of the superposition of the restrained energy minimized structures are shown in Figure 4, and the superposition of the average restrained dynamics structures is shown in Figure 5. The rms differences between all atoms, the sugar–phosphate backbone atoms, and the base atoms as a function of residue number are illustrated in Figure 6 for some combinations of structures.

Table III: Atomic rms Differences between Initial (IniI, IniII), Average Free Dynamics (FDI, FDII), Restrained Energy Minimized (RMI, RMII), and Average Restrained Dynamics (RDI, RDII) Structures

	overall rms difference (Å)						
	IniII	FDI	FDII	RMI	RMII	RDI	RDII
IniI	3.2	1.5	4.0	1.1	3.1	2.4	2.4
IniII		3.5	1.7	3.0	1.5	2.1	2.1
FDI			4.0	1.3	3.2	2.1	2.1
FDII				3.7	2.0	2.5	2.5
RMI					2.7	1.7	1.7
RMII						1.6	1.6
RDI							0.3

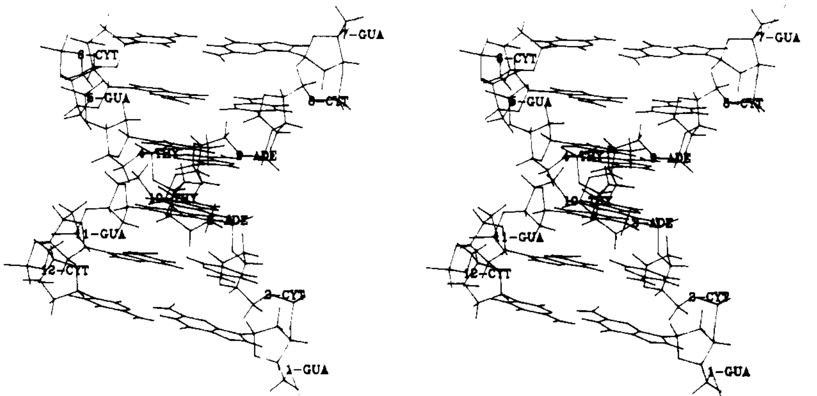
Table IV: rms Differences of the Interproton Distances for Initial (IniI, IniII), Average Free Dynamics (FDI, FDII), Restrained Energy Minimized (RMI, RMII), and Average Restrained Dynamics (RDI, RDII) Structures[a]

	rms difference of interproton distances (Å)		
	all (158)	intraresidue (96)	interresidue (62)
IniI	0.53	0.30	0.74
IniII	0.89	0.73	1.10
FDI	0.65	0.36	0.91
FDII	1.16	0.75	0.56
RMI	0.39	0.26	0.52
RMII	0.38	0.27	0.49
RDI	0.32	0.24	0.42
RDII	0.32	0.24	0.42

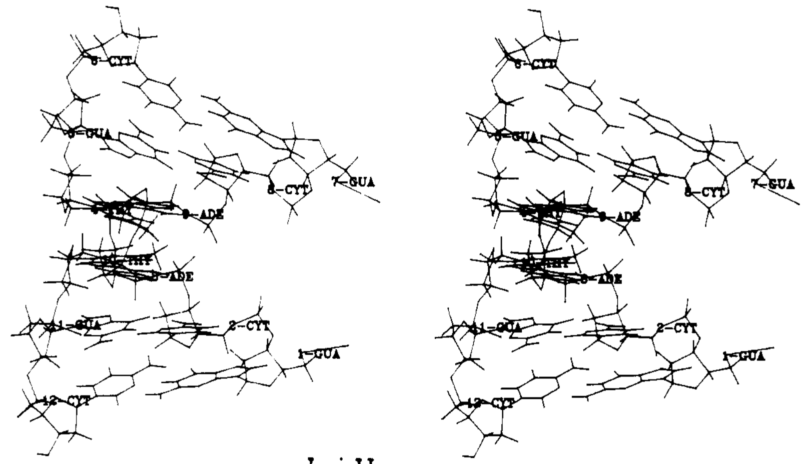
[a] Note that the total number of interproton distances is double the experimental number as the hexamer is self-complementary.

The rms difference of the interproton distances for both initial structures is >0.5 Å, and there are large numbers of distance violations >0.8 Å (28 and 58 for IniI and IniII, respectively). The free dynamics simulations explore a region of conformational space in the neighborhood of the starting structures. Thus, both average free dynamics structures are similar to their parent structures, although the atomic rms difference between them (4.0 Å) is a little larger than that between the two initial structures (3.2 Å). In addition, the rms difference of the interproton distances is increased for both average free dynamics structures relative to that for their respective starting structures (Table IV).

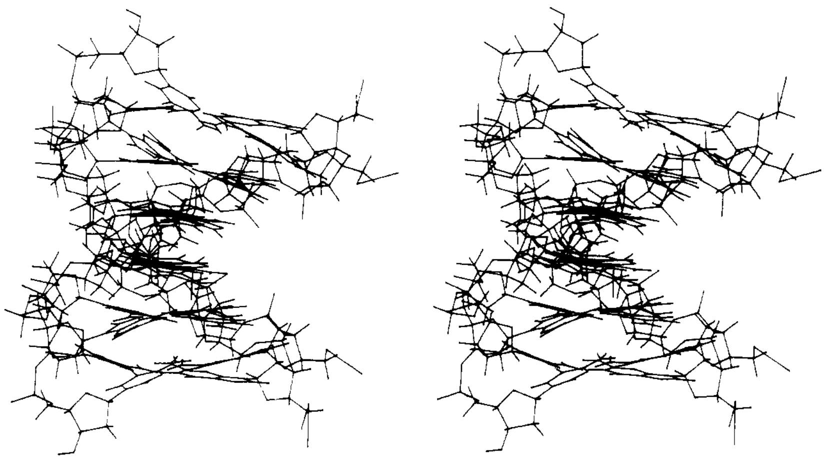
Both restrained energy minimization and restrained molecular dynamics reduce the overall rms difference of the interproton distances to <0.4 Å (Table IV). In this respect, restrained molecular dynamics is more effective than restrained



Ini I



Ini II



RMI vs RMII

FIGURE 4: Stereoviews along the helix axis of the initial (IniI, IniII) structures and the best fit superposition of the restrained energy minimized (RMI, RMII) structures. IniI and IniII are classical B- and A-DNA, respectively (Arnott & Hukins, 1972).

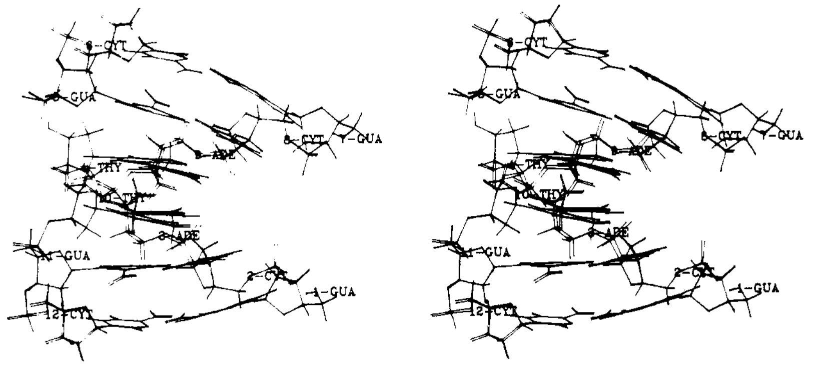
energy minimization not only in terms of the rms difference of the interproton distances (0.32 Å compared to 0.38–0.39 Å) but also with regards to the number of distance violations. The average restrained dynamics structure exhibits no distance violations >0.8 Å whereas the restrained energy minimized structures RMI and RMII have three and two distance violations >0.8 Å, respectively.

Despite the reduction in the rms difference of the interproton distances, restrained energy minimization fails to result in convergence. The atomic rms difference between the two restrained energy minimized structures (2.7 Å) is only a little smaller than that between the two initial structures. Indeed, although both restrained energy minimized structures have B-type glycosidic bonds and sugar pucker conformations, they retain the same global structure as that of their respective parent structures. In contrast to restrained energy minimization, restrained molecular dynamics results in convergence, both globally and locally, to essentially identical structures. The atomic rms difference between the two average restrained dynamics structures is only 0.3 Å. This is well within the rms fluctuations of the atoms about their average positions as can be ascertained from a comparison of the best

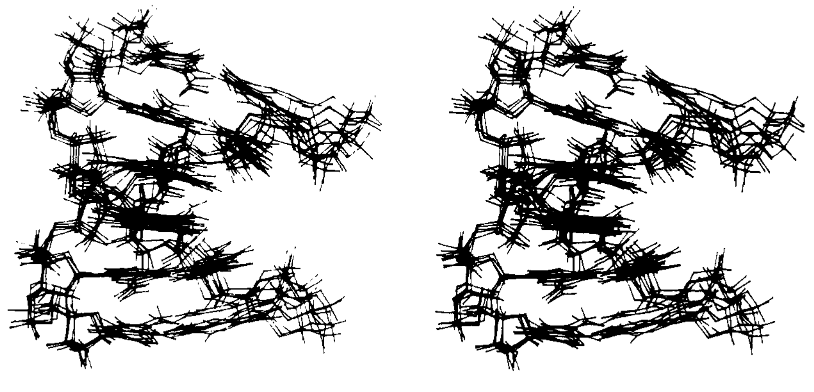
Table V: Individual Energy Terms for Initial (IniI, IniII), Average Free Dynamics (FDI, FDII), Restrained Energy Minimized (RMI, RMII), and Average Restrained Dynamics (RDI, RDII) Structures

	energy (kcal/mol) (number of terms)								
structures	total[a]	potential[a]	bond (406)	angle (732)	improper (172)	torsion (344)	electrostatic	van der Waals	hydrogen bonding (16)	restraints[b] (158)
IniI	391	69	15	147	0.04	240	−215	−77	−41	322
IniII	1318	322	129	227	0.06	214	−197	−20	−37	996
FDI[c]	[39][d]	−376	5	104	5	141	−284	−190	−57	[415][d]
FDII[c]	[1189][d]	−268	5	105	4	140	−286	−186	−50	[1457][d]
RMI	−64	−253	9	138	10	187	−257	−182	−58	189
RMII	−108	−304	7	152	11	203	−244	−171	−52	196
RDI[c]	−212	−320	8	101	6	159	−257	−178	−59	108
RDII[c]	−212	−321	8	101	7	161	−257	−179	−62	109

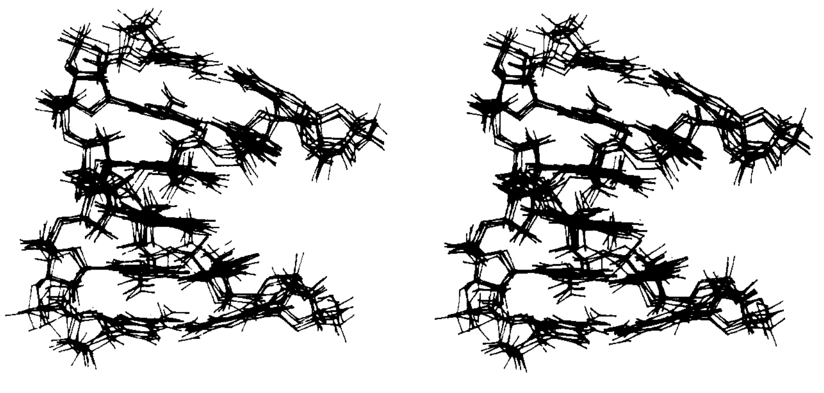
[a] The total energy includes the restraints energy whereas the potential energy does not. [b] The restraints scale factor S in eq 2 used for the restraints is 3. Thus, error estimates in the interproton distances of 0.2, 0.3, and 0.4 Å correspond to force constants of 22.4, 9.9, and 5.6 kcal/mol, respectively. [c] For the average free and restrained dynamics structures the energies are those obtained after subjecting the average structures to 500 cycles of energy minimization constrained to their original structures by weak harmonic constraints (Bruccoleri & Karplus, 1986). This procedure is used to correct for minor distortions in bond lengths and angles produced by the averaging procedure and results in only very small atomic rms shifts (<0.1 Å). [d] The total and restraint energies for the average free dynamics structures are shown in brackets as the restraints were not included at any stage of the calculations.



RDI vs RDII



RDI 10,15,20,25,30 ps



RDII 10,15,20,25,30 ps

FIGURE 5: Best fit superposition of (i) the two average restrained dynamics structures RDI and RDII and (ii) the structures at 10, 15, 20, 25, and 30 ps of the second dynamics run for the restrained dynamics structures RDI and RDII.

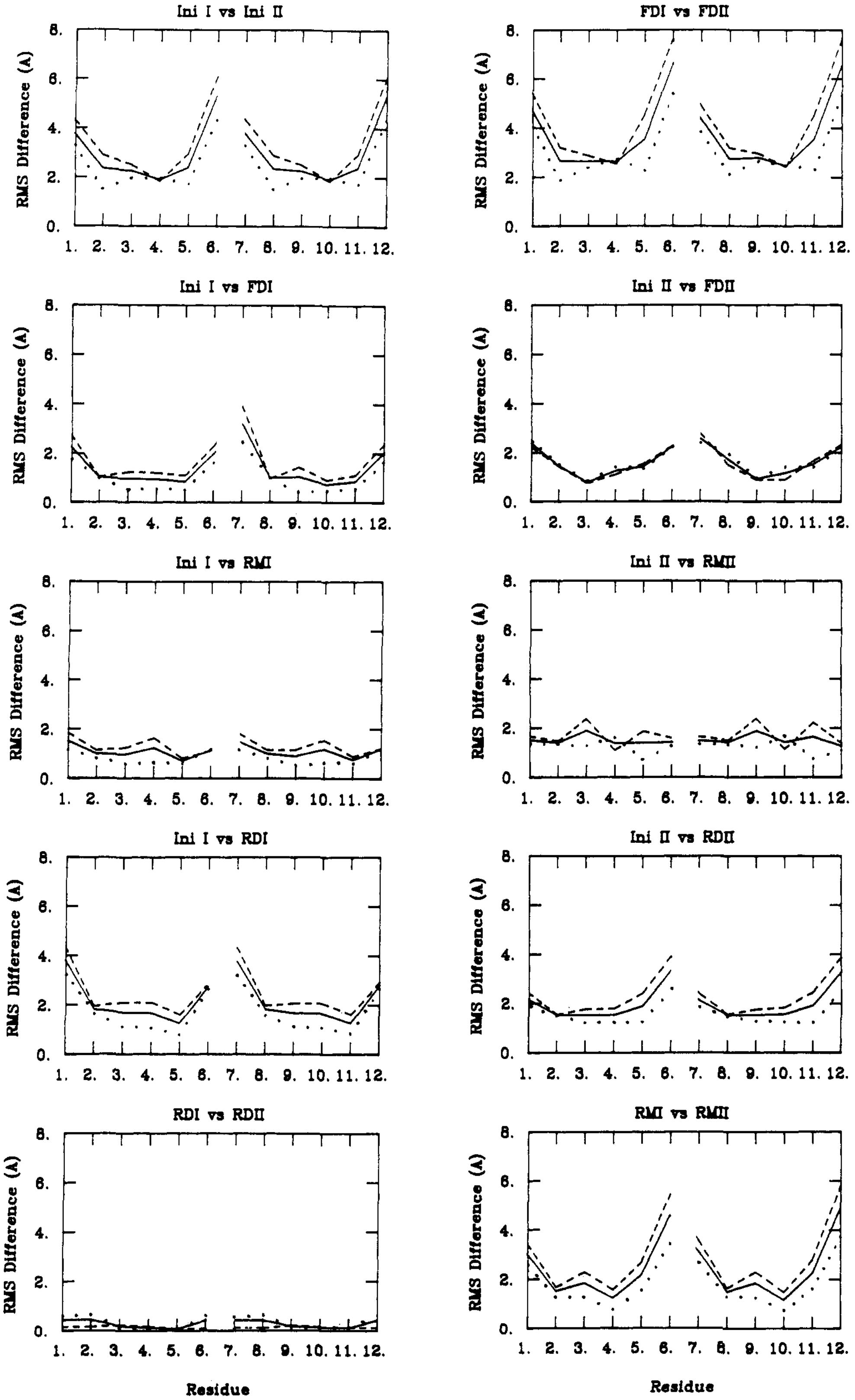


FIGURE 6: rms differences (Å) for all (—), the sugar–phosphate backbone (---), and the base (···) atoms as a function of residue number for various pairs of structures involving the initial (IniI, IniII), the restrained energy minimized (RMI, RMII), the average free dynamics (FDI, FDII), and the average restrained dynamics (RDI, RDII) structures.

fit superposition of the two average restrained dynamics structures with the best fit superpositions of snapshots taken at 10, 15, 20, 25, and 30 ps of the second dynamics run (Figure 5). The extent of convergence can also be gauged by the plots of torsion angles and helical parameters as a function of residue number for both average restrained dynamics structures shown in Figures 7 and 8.

A comparison of the energies of the average free and restrained dynamics structures as well as of the restrained energy minimized structures shows that the bond, angle, and improper energies are similar for the six structures, indicating that the restraining procedure has not resulted in any distortion of the covalent structure (Table V). The main difference in the energies of the average restrained dynamics structures and the

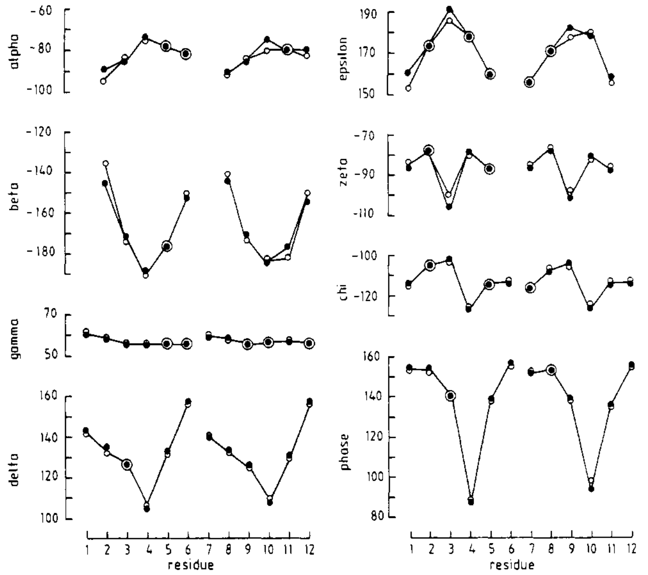


FIGURE 7: Variation in the backbone and glycosidic bond torsion angles as well as the phase angle describing the sugar pucker for the two restrained dynamics structures RDI (●) and RDII (○). The phase angle is calculated as described by Cremer and Pople (1975) with the apex at atom 3 and C4′ = atom 0, C1′ = atom 1, and so on.

restrained energy minimized structures involves two terms: the restraints and torsion energies, which are 80–90 kcal/mol and 30–40 kcal/mol lower, respectively, in the restrained dynamics structures than in the restrained energy minimized structures. There is no significant difference, however, in the nonbonding (i.e., electrostatic, van der Waals, and hydrogen bonding) energy terms.

The principal driving force in determining the outcome of the restrained dynamics simulations and the conformations of the structures, both in this case and in the case of the hexamer of Nilsson et al. (1986) and the decamer presented in the following paper (Nilges et al., 1987), is the NOE restraints energy, and the role of the dynamics is as a tool to overcome false minima and locate the global minimum region. At the same time the empirical energy function plays a role in so far that it maintains good stereochemistry and ensures good nonbonded contacts within the limitations of conformational space imposed by the NOE restraints. In conformational terms, the contribution of the empirical energy function is subsidiary to the NOE restraints energy, and further, the dominant terms of the empirical energy contribution appear to be different from case to case. Restrained energy minimization fails to result in convergence because it cannot overcome energy barriers and simply forces the initial structures into the closest local minimum. Free dynamics also does not result in convergence. This is not only because of defects in the empirical energy function (e.g., lack of solvent) but also, and more importantly, because of the absence of any large driving forces that would enable it to locate the global minimum region. That is to say, the differences in energies of the various false minima in the absence of the NOE restraints energy are so small that no matter how good the empirical energy function, free dynamics would always fail to achieve convergence. The absence of solvent in the restrained molecular dynamics calculations is unlikely to have any significant effect with respect to the outcome. The reason for this is that the NOE restraints energy terms themselves include the effect of solvent as the interproton distances are a direct measure of the actual solution structure under the experimental conditions employed. Finally, it is also worth pointing out that, in the case of proteins at least, the application of restrained molecular dynamics to structures generated by metric matrix distance geometry calculations results in an increase rather than a decrease in the atomic rms differences between the structures, while at the same time improving the structures energetically not only with respect to the nonbonding energies but also with respect to the NOE restraints energy (Clore et al., 1986b, 1987a,b).

Structural Features of the Average Restrained Dynamics Structures. The high degree of convergence of the two restrained dynamics simulations suggests that the two essentially identical average restrained dynamics structures RDI and RDII represent a reasonable approximation of the actual average structure in solution. Consequently, it is of interest to analyze these structures.

For all residues of the average restrained dynamics structures the backbone torsion angles lie in the conformational range expected for B-type DNA (Figure 7). The variation in the torsion angles exhibits almost perfect symmetry. This is not entirely unexpected given the symmetry of the distance restraints and the symmetry of the starting structures (but see below). It should be noted, however, that there are deviations from perfect symmetry. This arises because once asymmetry is introduced during the course of the restrained dynamics simulation, for example, as a result of the asymmetric assignment of initial velocities with respect to the axis of symmetry, it will persist. This is in marked contrast to the relatively large deviations in symmetry observed in the crystal structure of the self-complementary B-DNA dodecamer (Dickerson & Drew, 1981). It is important in this regard to note that, because of the self-complementary nature of the hexamer, as well as that of the decamer presented in the following paper (Nilges et al., 1987), the effective number of restrained dynamics calculations is double the actual number. The reason for this lies in the fact that the initial velocities are not assigned symmetrically with respect to the axis of symmetry so that the convergence pathways followed by the two symmetrical halves of the oligonucleotide are different. Indeed, experience with model calculations on crambin indicates that the assignment of initial velocities has a substantial effect on the final structures. Thus, in the case of the crambin, calculations starting from a completely extended strand and using protocols identical in all features except for the value of the random number seed used to assign the initial velocities result in a set of final converged crambin-like structures that are as different from each other (A. T. Brunger, G. M. Clore, A. M. Gronenborn, and M. Karplus, unpublished results) as those obtained by Clore et al. (1986a) and Brunger et al. (1986) using different starting structures and different protocols. Consequently, the departures from perfect symmetry seen in the restrained dynamics oligonucleotide structures are of quite a different nature and are much larger than the very small deviations due to rounding errors seen in restrained energy minimization and restrained least-squares refinement calculations.

Another feature of interest in Figure 7 is the good agreement in the values of α, β, γ, ϵ, and ξ angles for the two refined structures, RDI and RDII, despite the fact that these backbone torsion angles are not directly related to any of the interproton distances measured. These values, however, are not coincident in the two free dynamics structures, FDI and FDII. It would therefore appear that the positioning of each nucleotide unit relative to each other achieved by the NOE restraints is sufficient, in the presence of the empirical energy function, to localize these backbone torsion angles to relatively narrow regions of conformational space within the range of values that

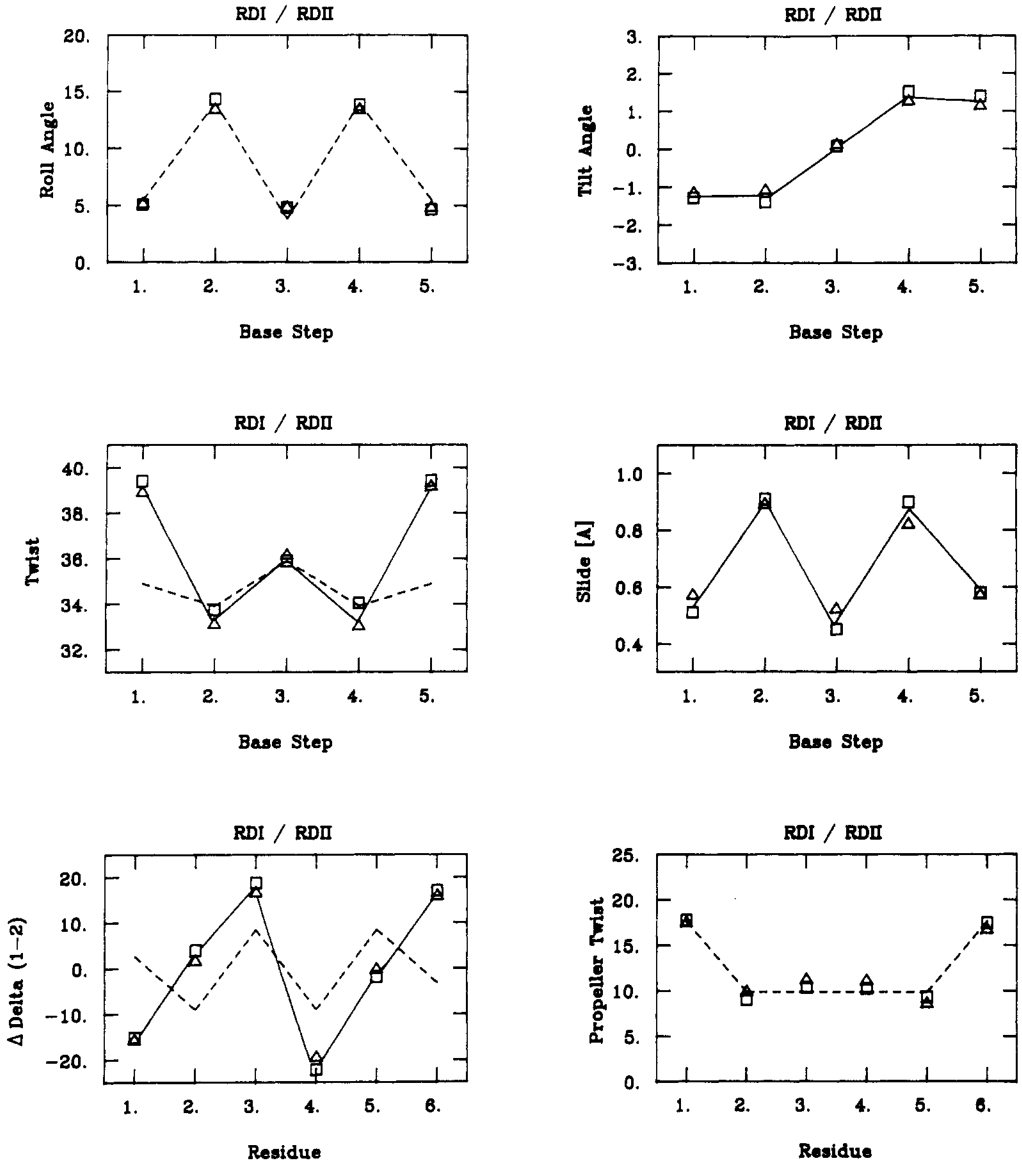


FIGURE 8: Variation in base roll, base tilt, global helical twist, and propeller twist angles, in slide, and in the difference of the C4′–C3′ bond torsion angles δ at the two ends of a base pair ($\Delta\delta_{1-2}$), for the average restrained dynamics structures RDI (□) and RDII (△). The dashed lines represent the best fits of Dickerson's (1983) sum functions ($\sum_1$–$\sum_4$) to the data. $\sum_1$ is the sum function for global helical twist, $\sum_2$ for base roll, $\sum_3$ for $\Delta\delta_{1-2}$, and $\sum_4$ for propeller twist. The best fits are calculated by using the equation $y = S + T\sum$, where y is the experimental value. The terms for $\sum_1$ are +1, −2, and +1 for x-Pur-Pyr-x and +2, −4, and +2 for x-Pyr-Pur-x; for $\sum_2$, they are +1, −2, and +1 for x-Pur-Pyr-x and −2, +4, and −2 for x-Pyr-Pur-x; for $\sum_3$, they are +1 and −1 for Pur-Pyr and −2 and +2 for Pyr-Pur; and for $\sum_4$, they are −1 and −1 for Pur-Pyr and −2 and −2 for Pyr-Pur. In the case of global helical twist and $\Delta\delta_{1-2}$, the fit shown represents the fit to the central four base pairs only.

can be adopted by A- and B-DNA structures (viz., anywhere in the g⁻, t, g⁺, t, and g⁻ ranges for the α, β, γ, ϵ, and ξ angles, respectively).

In order to analyze the local helical properties of the restrained dynamics structures, local helix rotation and translation vectors were defined by using the vectors between the two C1′ atoms and their attached base nitrogen atoms of one base pair and the corresponding atoms of the next base pair. The local helix vectors are shown in Figure 9 on a skeletal drawing of the restrained dynamics structure RDI, where it can be seen that they are not aligned along the long axis of the hexamer. Rather, each local helix vector is displaced and tilted with respect to the vectors at neighboring steps. The local helix vectors themselves can be classified into two sets: namely, the set comprising the Pur-Pyr base steps 1, 3, and 5 and that comprising the Pyr-Pur base steps 2 and 4. The former describe a bend to the left while the latter describe a bend to the right, in the plane of Figure 9a.

The GpC base steps 1 and 5 have values of local helical parameters typical for B-DNA. The base planes are almost perpendicular to the local helix axis and are not much displaced from it. The local helical twist is a little larger than average (37.6–38.0° corresponding to 9.6–9.7 base pairs per turn), a finding in agreement with the observation of higher than average local helical twists for the GpC steps of the B-DNA dodecamer (Dickerson & Drew, 1981). In addition, the bases exhibit good intrastrand stacking, typical of B-DNA, with extensive overlap of the six-membered ring of the purine base with the pyrimidine ring of the adjacent 3′-residue (Figure 10).

The Pyr-Pur steps 2 and 4, on the other hand, display a number of features that bear semblance to an A-type helix. The rotation axis is displaced significantly into the major groove, and the base pairs are tilted clockwise relative to the local helix axis by 31° when viewed into the minor groove. This variation in local helix type is the same as that found for the nonterminal Pyr-Pur steps in the B-DNA dodecamer

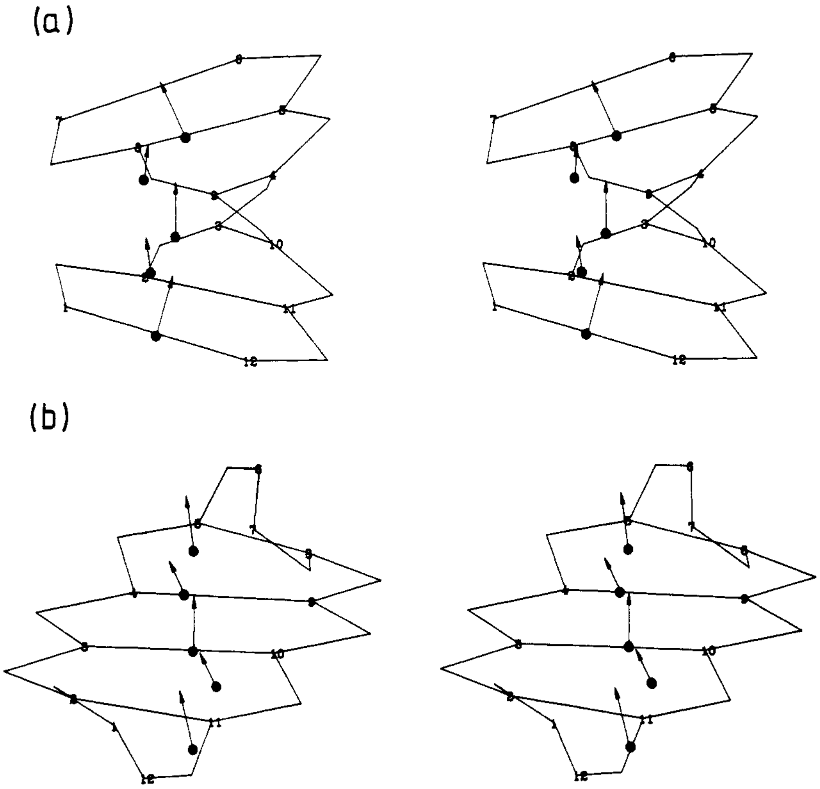


FIGURE 9: Two stereoviews of the average restrained dynamics structure RDI with individual local helix rotation vectors relating two successive base pairs superimposed. Only the phosphorus and C1′ atoms are retained with the residue labels at the C1′ atoms. The two views are separated by a 90° rotation along the long axis of the molecule.

(Dickerson & Drew, 1981). In contrast to the GpC steps, the two Pyr-Pur steps exhibit no intrastrand stacking and a small degree of interstrand stacking with partial overlap of the six-membered ring of the purines (Figure 10). This too is rather typical of an A-type helix [see, for example, Figure 5 of Shakked et al. (1983)].

The central ApT base step has the lowest value for the local helical twist (34.2–34.8°). An even lower value is found for the ApT step of the B-DNA dodecamer (31.2°; Dickerson & Drew, 1981). In the A-DNA octamer, on the other hand, the local helical twist is largest at the central ApT step with a value of 34.1° which is almost identical with that found in the present structure. Interestingly, a similar value (34.3°) has been measured by Rhodes and Klug (1981) for poly[d(AT)] in solution. In other respects, such as displacement and base stacking, the ApT step has B-like features. Note again the extensive intrastrand overlap of the six-membered ring of the purine with the pyrimidine ring (Figure 10).

A striking feature of the average restrained dynamics structures is that they appear bent (see Figures 5 and 9). The extent of bending can be estimated by calculating the angle between the local helix axis vectors of similar steps (i.e., only Pur-Pyr or only Pyr-Pur steps). The angle between the local helix vectors of steps 1 and 3 is 24.5°, and that between those of steps 3 and 5 is 25.7°. Most of the bending occurs at the two intervening Pyr-Pur steps 2 and 4 which, as discussed in the previous sections, resemble an A-type helix. In this respect, we note that most of the bending in the bent variants of the B-DNA dodecamer also occurred near Pyr-Pur steps (Fratini et al., 1982). As the two Pyr-Pur steps are only separated by one base pair in the hexamer, both bends are almost in the same direction, and the total angle between the local helix vectors of steps 1 and 5 is 47.6°. This corresponds to a radius of curvature of approximately 20 Å. Given that there are no external forces present in solution that could induce bending, in contrast to the situation in the crystal state where packing forces may play an important role, it is clear that the bending of the hexamer arises as a consequence of base sequence.

At first it seems surprising that the present bent structure should possess a best overall helix axis, though from symmetry considerations it is clear that there has to be one. None of the local helix axis vectors coincide with this best overall helix axis, which we will term the "superhelix" axis. Rather, the local helix axis vectors wind around the superhelix axis in two separate helices: the local helix axis vectors at the Pur-Pyr B-like steps in one and those of the Pyr-Pur A-like steps in the other. This is illustrated in Figure 11, which shows an "infinite" helix GCATGCATGC... derived from the coordinates of the average restrained dynamics structure RDI by repeatedly rotating and translating an image of the original helix until the first two base pairs of the shifted image overlap and coincide with the last two base pairs of the original. The resulting structure is compressed relative to both B- and A-DNA: as a result, it is shorter and fatter than either of the two idealized structures. In addition, the distinction between the major and minor groove is essentially lost. The width and depth of the major groove are reduced whereas those of the minor groove are increased relative to classical B-DNA. Thus the appearance when viewed along the long axis is that of a very narrow waist repeated at regular intervals along a rather fat body (Figure 11a). When viewed down the helix axis (Figure 11b), the structure appears intermediate between A- and B-DNA although somewhat closer to the latter than the former.

Sequence-dependent variations in local structure have been reasonably successfully described by Calladine's rules (Calladine, 1982) for a number of A- and B-DNA crystal structures (Fratini et al., 1982; Dickerson, 1983). According to these rules, a helix can reduce the steric clash between purine residues of opposite strands at Pur-Pyr and Pyr-Pur steps by

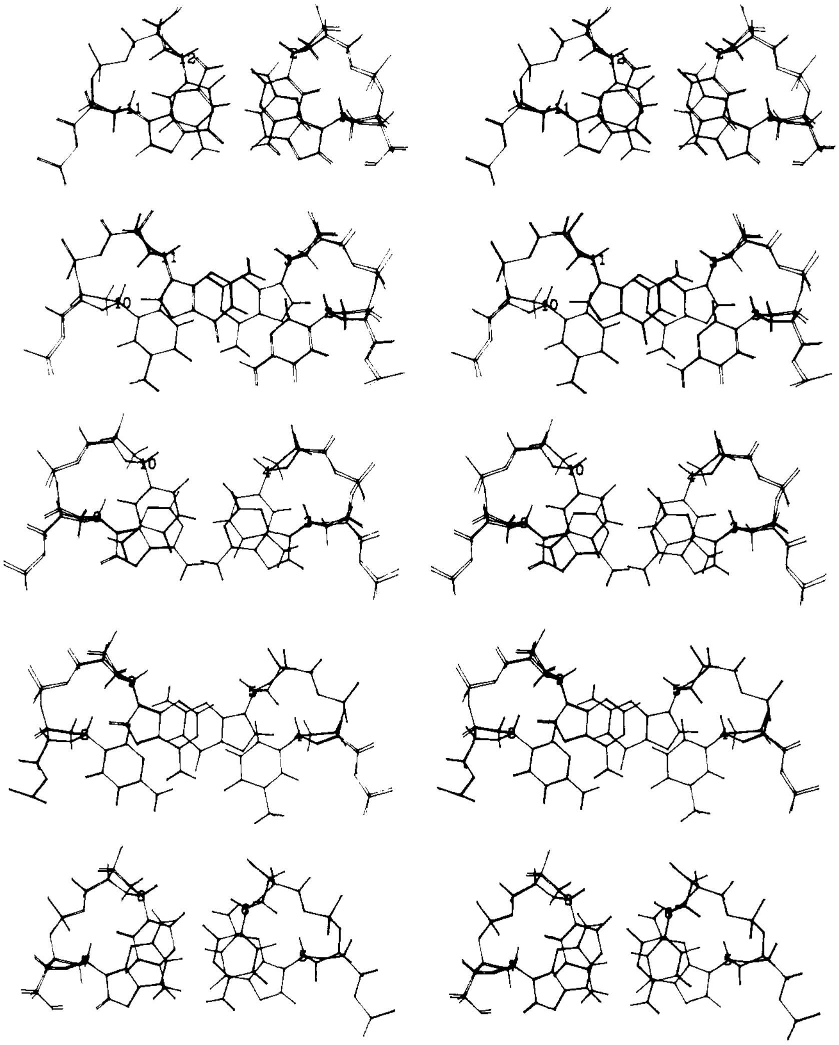


FIGURE 10: Stereoviews of the best fit superposition of the five individual base pair steps of the two average restrained dynamics structures, RDI and RDII, viewed down the helix axis.

means of four different strategies: (i) a decrease in helical twist at these steps; (ii) opening up of the base roll angle on the side of the clash; (iii) separating the purine residues by sliding the base pairs involved along their long axis; and (iv) flattening of the propeller twist in one or both base pairs. These four strategies have been put on a quantitative basis by means of a simple set of sum functions $\sum_1$, $\sum_2$, $\sum_3$, and $\sum_4$ which describe the global helical twist, the base roll angle, the difference in the C4′–C3′ bond torsion angles δ in a single base pair ($\Delta\delta_{1-2}$), and the propeller twist, respectively (Dickerson, 1983). The sum functions are then linearly fitted to the experimental data to yield a regression line $y = S + T\sum$, where y is the experimental value. The variation in helical parameters for the average restrained dynamics structures, RDI and RDII, is shown in Figure 9 together with the sum function predictions.

The global helical twist shows an alternating pattern with higher than average values at Pur-Pyr steps and lower than average values at Pyr-Pur steps (Figure 8). The average values of the global (36.3 ± 3) and local helical twist (36.6 ± 1.4), however, are essentially identical. A linear fit of the global helical twist to Calladine's rules only gives a significant correlation if the end base pairs are neglected. The values of the best fit parameters thus obtained are $S = 34.9°$ and $T = 0.5°$ with the correlation coefficient $R = 0.993$. This compares to values of 35.6° and 2.1°, respectively, for the B-DNA dodecamer (Dickerson, 1983).

While base tilt shows little sequence-dependent variation, the base roll angle alternates strongly with sequence. Pur-Pyr and Pyr-Pur steps open toward the minor groove with roll angles of 5° and 14°, respectively. The excellent agreement with Calladine's rules (correlation coefficient $R = 0.993$, $S = 8.9°$ and $T = 0.85°$) suggests that the purine–purine clash in the minor groove at the Pyr-Pur steps is responsible for the variation in base roll. Unusual for B-DNA, however, is the fact that all the base steps open toward the minor groove. The mean of the base roll angles is shifted to over 8° compared to a value of −2.3° in the B-DNA dodecamer (Dickerson & Drew, 1981). As a result, the large and positive values for the base roll angles at the Pyr-Pur steps are not counterbalanced by negative values at the Pur-Pyr. This is the underlying structural feature responsible for the observed bending of the average restrained dynamics structures. One can speculate that one of the reasons for the shift in base roll angles to entirely positive values is the absence of homopolymer steps in the sequence (i.e., Pur-Pur or Pyr-Pyr) which could resist roll in either direction (Dickerson & Drew, 1981).

In a similar fashion to global helical twist and base roll, the base pair slide also alternates with base sequence: it has values around 0.5 and 1 Å for the Pur-Pyr and Pyr-Pur steps, respectively (Figure 8). The combined effect of base roll and slide on the average restrained dynamics structures can be understood in the following manner. Calladine and Drew (1984) noted that an A helix can be generated from a B helix

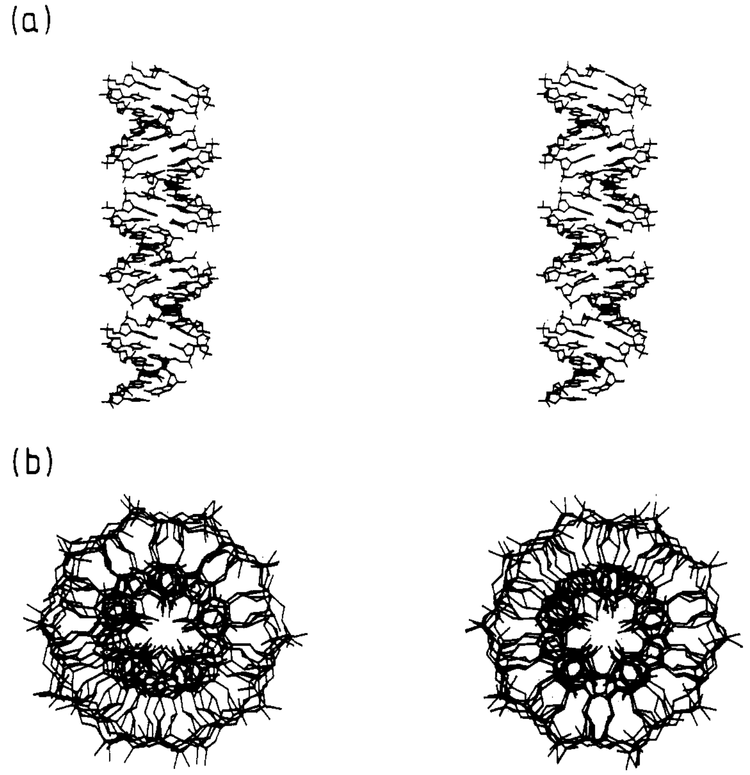


FIGURE 11: Stereoviews of the 26 base pair long helices of sequence GCATGCATGCATGCATGCATGCATGC generated from the coordinates of the average restrained dynamics structure RDI viewed along (a) and down (b) the helix axis. This structure was derived by repeatedly rotating and translating an image of the original six base pair helix until the first two base pairs of the shifted image overlap and coincide with the last two base pairs of the original.

by introducing positive roll and slide. In the average restrained dynamics structures roll is large and positive, and slide is larger than in classical B-DNA although smaller than in classical A-DNA. The variation in slide is about 0.5 Å from step to step. Similar values have been found in all crystal structures to date, and it has been proposed on this basis that slide may communicate the overall structure from one base pair to the next (Calladine & Drew, 1984). In order to compare our data for the average restrained dynamics structures with that for the crystal structures, we have depicted roll and slide in a roll–slide diagram (Figure 12). All the Pur-Pyr steps are centered around a point (slide ~0.5 Å, roll ~5°) in the B-type region close to the border between A- and B-type geometries. The Pyr-Pur steps, on the other hand, are centered around a point (slide ~1 Å, roll ~14°) that is in the A-type region but is left shifted with respect to classical A-DNA. The shaded areas in the diagram indicate the mobility of the base pairs with the width and height approximately equal to the rms fluctuations in slide and roll, respectively.

The third of Calladine's rules is not directly applicable to slide but to the related difference in the C4′–C3′ bond torsion angle δ at the two ends of a base pair, $\Delta\delta_{1-2}$ (Calladine, 1982; Fratini et al., 1982; Dickerson, 1983). The correlation between the observed values of $\Delta\delta_{1-2}$ and the sum function prediction, however, is rather poor (Figure 8). The reason for this is that as the average restrained dynamics structures are bent, the C4′–C3′ bond torsion angle δ is not parallel to the helix axis so that the pulling of a purine out of the helix stack is not as closely related to the value of δ as in a straight B-DNA helix.

Propeller twist is flattened in the central four base pairs as predicted by Calladine's fourth rule, and the correlation with the sum function prediction is good ($R = 0.993$). The fit parameters S and T have values of 21.6° and 3.9°, respectively, which are close to those found for the B-DNA dodecamer (24.3° and 3.6°, respectively; Dickerson, 1983).

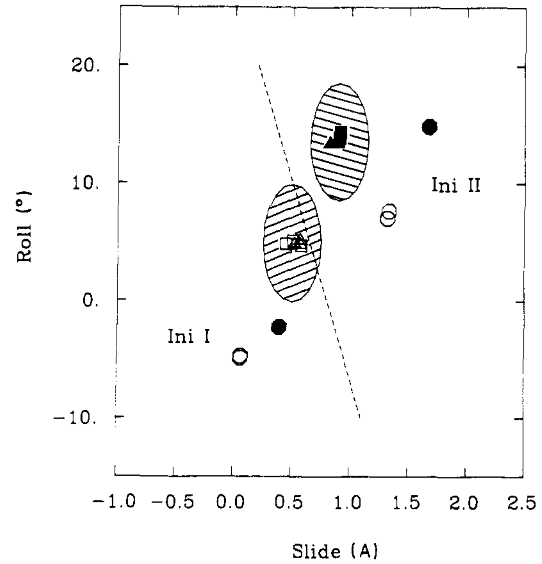


FIGURE 12: Roll–slide diagram for the average restrained and initial structures. Symbols: □ and ■, RDI; △ and ▲, RDII; ○ and ●, IniI (classical B-DNA) and IniII (classical A-DNA). Open and closed symbols represent Pur-Pyr and Pyr-Pur steps, respectively. The dashed line stretching from roll, slide, = −10°, 1Å, to +20°, 0.2 Å, represents the break between A and B geometries, which lie to the right and left, respectively, of the line (Calladine & Drew, 1984).

Thus, in the case of the average restrained dynamics structures, the flattening of propeller twist and the increase in base roll angle at Pyr-Pur steps appear to be the main strategies for relieving interstrand purine–purine clash, judging from the very good correlation between the data and Calladine's rules 2 and 4.

CONCLUSIONS

In this paper we have used NOE measurements combined with restrained molecular dynamics to refine the structure of the self-complementary hexamer 5′d(GCATGC)$_2$ in solution. The restrained dynamics simulations starting from classical A- and B-DNA converge to essentially identical structures with an atomic rms difference (0.3 Å) within the rms fluctuations of the atoms about their average positions, suggesting that they represent reasonable approximations to the "actual" average solution structure. The average restrained dynamics structures exhibit clear sequence-dependent structural variations, particularly in the helical parameters, some of which are in accord with Calladine's rules. The restrained dynamics structures are also bent, a structural feature that can be almost entirely attributed to the large positive base roll angles, particularly at the Pyr-Pur steps.

The structure of the hexamer 5′d(GCATGC)$_2$ presented in this paper and that of the decamer 5′d(CTGGATCCAG)$_2$ presented in the following paper (Nilges et al., 1987) probably represent the first or at least the most dramatic examples of the determination of the structures of extended nonglobular molecules for which long-range (i.e., global) structural conclusions are drawn from short range (<5-Å) interproton distance data. The structural features that emerge are entirely a result of the incorporation of the NOE interproton distances into the total energy function of the system in the form of effective potentials and are *not* in any way artifacts arising from the empirical energy function. The convergence to unique structures is due to the power of restrained molecular dynamics as a tool with which to overcome local energy barriers and escape false minima, thereby locating the global minimum region. Thus the NOE restraints energy guides the refinement to the correct global minimum region, and the empirical energy function ensures that the local stereochemistry and nonbonded interactions are approximately correct. In this respect it is essential that the force constants for the NOE restraints are sufficiently high to ensure that all the interproton distances are satisfied within the errors specified while taking care that this does not result in bad nonbonded contacts. In the absence of the NOE restraints *no* convergence occurs. At the same time, it should be noted that many of the structural features, including the bending, cannot be readily attributed to individual restraints. The reason for this is that virtually all the measurable interproton distances within any block of three consecutive base pairs are highly correlated to all the structural parameters. Consequently, it would not have been possible to deduce any of the structural features, with the possible exception of the glycosidic bond and sugar pucker conformations, from a qualitative interpretation of the NOE data (Gronenborn & Clore, 1985). It is precisely because of this that it is absolutely essential in studies of this kind to start the structure determination or refinement from significantly different initial structures (e.g., in the case of DNA, A and B types). If convergence occurs to a unique structural set that satisfies the interproton distance restraints, then one can be confident that a realistic and accurate picture of the solution structure has been obtained and that the global minimum region has been located. Conversely, if convergence does not occur, then one can conclude that the experimental data are insufficient to determine the structure.

At the present time it is not clear what the significance of this structural feature may be in relation to the role of the hexamer as the specific target site for the restriction endonuclease *Sph*I. We note, however, that *Sph*I cleaves the hexamer at the G$_1$pC$_2$ step and the symmetrically related G$_7$pC$_8$ step so that the bending, if it occurs in the *Sph*I–DNA complex, may serve to bring these two sites closer together. That such bending may indeed be important in DNA–protein interactions is further supported by the direct observation of DNA bending in the specific DNA–*Eco*RI restriction endonuclease complex by X-ray crystallography (Frederick et al., 1984) and in the DNA–cAMP receptor protein complex by electron microscopy (Gronenborn et al., 1984b). DNA bending has also been demonstrated indirectly by the observation of altered electrophoretic mobility for specific DNA–cAMP receptor protein (Wu & Crothers, 1984) and DNA–SV40 large T antigen (Ryder et al., 1986) complexes. Indeed, in the latter case, it was shown that specific binding only occurs on DNA that is already bent in the free state (Ryder et al., 1986). In addition, DNA bending has been shown to be important in determining nucleosome positioning (Drew & Travers, 1985). When the restrained dynamics structures are extended to "infinite" helices, it is seen that a superhelix is present. These "infinite" helices are compressed relative to both B- and A-DNA, and such structures may be of relevance to DNA packaging in the absence of chromatin.

ACKNOWLEDGMENTS

We thank the Max-Planck-Institut für Plasma Physik (Garching) for computing facilities on the CRAY 1 computer.

REFERENCES

Arnott, S., & Hukins, D. W. L. (1972) *Biochem. Biophys. Res. Commun. 47*, 1504–1509.

Bax, A., & Davis, D. G. (1985) *J. Magn. Reson. 65*, 355–360.

Bodenhausen, G., Vold, R. L., & Vold, R. R. (1980) *J. Magn. Reson. 37*, 93–106.

Bolton, P. H., & James, T. L. (1979) *J. Phys. Chem. 83*, 3359–3367.

Bolton, P. H., & James, T. L. (1980) *J. Am. Chem. Soc. 102*, 25–31.

Braun, W., & Go, N. (1985) *J. Mol. Biol. 186*, 611–626.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187–217.

Bruccoleri, R. E., & Karplus, M. (1986) *J. Comput. Chem. 7*, 175–185.

Brunger, A. T., Clore, G. M., Gronenborn, A. M., & Karplus, M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3801–3805.

Calladine, C. R. (1982) *J. Mol. Biol. 161*, 343–352.

Calladine, C. R., & Drew, H. R. (1984) *J. Mol. Biol. 178*, 773–782.

Clore, G. M., & Gronenborn, A. M. (1983) *EMBO J. 2*, 2109–2115.

Clore, G. M., & Gronenborn, A. M. (1984) *FEBS Lett. 172*, 219–225.

Clore, G. M., & Gronenborn, A. M. (1985a) *FEBS Lett. 179*, 187–198.

Clore, G. M., & Gronenborn, A. M. (1985b) *J. Magn. Reson. 61*, 158–164.

Clore, G. M., Lauble, H., Frenkiel, T. A., & Gronenborn, A. M. (1984) *Eur. J. Biochem. 145*, 629–636.

Clore, G. M., Gronenborn, A. M., Brunger, A. T., & Karplus, M. (1985) *J. Mol. Biol. 186*, 435–455.

Clore, G. M., Brunger, A. T., Karplus, M., & Gronenborn, A. M. (1986a) *J. Mol. Biol. 191*, 523–551.

Clore, G. M., Nilges, M., Sukumaran, D. K., Brunger, A. T., Karplus, M., & Gronenborn, A. M. (1986b) *EMBO J. 5*, 2729–2735.

Clore, G. M., Sukumaran, D. K., Nilges, M., & Gronenborn, A. M. (1987a) *Biochemistry 26*, 1732–1745.

Clore, G. M., Sukumaran, D. K., Nilges, M., Zarbock, J., & Gronenborn, A. M. (1987b) *EMBO J. 6*, 529–537.

Cremer, D., & Pople, J. A. (1975) *J. Am. Chem. Soc. 97*, 1358–1367.

Crippen, G. M., & Havel, T. F. (1978) *Acta Crystallogr., Sect. A*: *Cryst. Phys., Diffr., Theor. Gen. Crystallogr. A34*, 282–284.

Davis, D. G., & Bax, A. (1985) *J. Am. Chem. Soc. 107*, 2821–2822.

Dickerson, R. E. (1983) *J. Mol. Biol. 166*, 419–441.

Dickerson, R. E., & Drew, H. R. (1981) *J. Mol. Biol. 149*, 761–786.

Dickerson, R. E., Drew, H. R., Conner, B. N., Kopka, M. L., & Pjura, R. E. (1983) *Cold Spring Harbor Symp. Quant. Biol. 47*, 13–24.

Dobson, C. M., Olejniczak, E. T., Poulsen, F. M., & Ratcliffe, R. G. (1982) *J. Magn. Reson. 48*, 87–110.

Drew, H. R., & Dickerson, R. E. (1981) *J. Mol. Biol. 151*, 535–556.

Drew, H. R., & Travers, A. A. (1984) *Cell* (*Cambridge, Mass.*) *77*, 491–502.

Drew, H. R., & Travers, A. A. (1985) *J. Mol. Biol. 186*, 773–790.

Feigon, J., Denny, W. A., Leupin, W., & Kearns, D. R. (1983) *Biochemistry 22*, 5930–5942.

Fratini, A. V., Kopka, M. L., Drew, H. R., & Dickerson, R. E. (1982) *J. Biol. Chem. 257*, 14686–14707.

Frederick, C. A., Grable, J., Melia, M., Sanudzi, C., Jen-Jacobsen, L., Wang, B. C., Greene, P., Boyer, H. W., & Rosenberg, J. M. (1984) *Nature* (*London*) *309*, 327–331.

Gelin, G. R., & Karplus, M. (1975) *Proc. Natl. Acad. Sci. U.S.A. 72*, 2002–2006.

Gronenborn, A. M., & Clore, G. M. (1985) *Prog. Nucl. Magn. Reson. Spectrosc. 17*, 1–33.

Gronenborn, A. M., Clore, G. M., & Kimber, B. J. (1984a) *Biochem. J. 221*, 723–736.

Gronenborn, A. M., Nermut, M. V., Easton, P., & Clore, G. M. (1984b) *J. Mol. Biol. 179*, 751–757.

Hare, D. R., & Reid, B. R. (1986) *Biochemistry 25*, 5341–5350.

Hare, D. R., Wemmer, D. E., Chou, S. H., Drobny, G., & Reid, B. R. (1983) *J. Mol. Biol. 171*, 319–336.

Havel, T. F., & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 281–294.

Hogan, M. E., & Jardetzky, O. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 6341–6345.

Jeener, J., Meier, B. H., Backmann, P., & Ernst, R. R. (1979) *J. Chem. Phys. 71*, 4546–4553.

Jones, T. A. (1978) *J. Appl. Crystallogr. 11*, 268–272.

Jones, T. A. (1982) in *Computational Crystallography* (Sayre, D., Ed.) pp 303–317, Clarendon, Oxford.

Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R., & van Gunsteren, W. F. (1985) *J. Mol. Biol. 182*, 179–182.

Keepers, J. W., & James, T. L. (1984) *J. Magn. Reson. 53*, 404–420.

Kline, A. D., Braun, W., & Wüthrich, K. (1986) *J. Mol. Biol. 189*, 377–382.

Lomonossoff, G. P., Butler, P. J. G., & Klug, A. (1981) *J. Mol. Biol. 149*, 745–760.

Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967–974.

Matteuci, M. D., & Caruthers, M. H. (1981) *J. Am. Chem. Soc. 103*, 3185–3191.

McCall, M. J., Brown, T., & Kennard, D. (1985) *J. Mol. Biol. 183*, 385–396.

Nilges, M., Clore, G. M., Gronenborn, A. M., Piel, N., & McLaughlin, L. W. (1987) *Biochemistry* (following paper in this issue).

Nilsson, L., & Karplus, M. (1986) *J. Comput. Chem. 7*, 691–716.

Nilsson, L., Clore, G. M., Gronenborn, A. M., Brunger, A. T., & Karplus, M. (1986) *J. Mol. Biol. 188*, 455–476.

Noggle, J. H., & Schirmer, R. E. (1971) *The Nuclear Overhauser Effect—Chemical Application*, Academic, New York.

Redfield, A. G., & Kunz, S. D. (1975) *J. Magn. Reson. 19*, 250–254.

Reid, D. G., Salisbury, S. A., Bellard, S., Shakked, Z., & Williams, D. H. (1983) *Biochemistry 22*, 2019–2025.

Rhodes, D. R. (1982) in *Topics in Nucleic Acids Structures, Part 2* (Neidle, S., Ed.) pp 287–304, MacMillen, London.

Rhodes, D., & Klug, A. (1980) *Nature* (*London*) *286*, 573–578.

Ryckaert, J. P., Cicotti, G., & Berendsen, H. J. C. (1977) *J. Comput. Phys. 23*, 327–337.

Ryder, K., Silver, S., Delucia, A. L., Fanning, L., & Tagtmeyer, P. (1986) *Cell* (*Cambridge, Mass.*) *44*, 719–725.

Scheek, R. M., Russo, N., Boelens, R., Kaptein, R., & van Boom, J. H. (1983) *J. Am. Chem. Soc. 105*, 2914–2916.

Seliger, H., Klein, S., Narang, C. K., Seeman-Preising, B., Eiband, J., & Haud, N. (1982) in *Chemical and Enzymatic Synthesis of Gene Fragments*: *A Laboratory Manual* (Gassen, G. H., & Lang, A., Eds.) pp 81–96, Verlag Chemie, Weinheim.

Shakked, Z., Rabinovich, D., Kennard, D., Cruse, W. B. T., Salisbury, S. A., & Viewamitra, A. (1983) *J. Mol. Biol. 166*, 183–201.

Solomon, I. (1955) *Phys. Rev. 99*, 559–565.

Tidor, B., Irikura, K., Brooks, B. R., & Karplus, M. (1983) *J. Biomol. Struct. Dyn. 1*, 231–252.

Verlet, L. (1967) *Phys. Rev. 159*, 98–105.

Wagner, G., & Wüthrich, K. (1979) *J. Magn. Reson. 33*, 675–680.

Wang, A. H.-J., Fujii, S., van Boom, J. H., & Rich, A. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 5470–5474.

Weiss, M. A., Patel, D. J., Sauer, R. T., & Karplus, M. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 130–134.

Williamson, M. P., Havel, T. F., & Wüthrich, K. (1985) *J. Mol. Biol. 182*, 295–315.

Wu, H., & Crothers, D. M. (1984) *Nature* (*London*) *308*, 509–513.